An evaluation of physical stream habitat attributes used to monitor streams.

Brett B. Roper
Fish and Aquatic Ecology Unit, US Forest Service
860 North 1200 East
Logan UT, 84321

Jeff L. Kershner
Fish and Aquatic Ecology Unit, US Forest Service
5210 University Hill
Logan, UT 84322-5210

Eric Archer
Fish and Aquatic Ecology Unit, US Forest Service
860 North 1200 East
Logan UT, 84321

Richard Henderson
Fish and Aquatic Ecology Unit, US Forest Service
860 North 1200 East
Logan UT, 84321

Nicolaas Bouwes
Eco Logical Research
456 South 100 West
Providence, UT 84332

Abstract

The last couple decades have seen an increased reliance on the use of stream attributes to monitor stream conditions. The use of stream attributes have been criticized because of variation in how observers evaluate them, inconsistent protocol application, lack of consistent training, and the difficultly in using them to detect change due to management activity. In this paper, we evaluate the effect of environmental heterogeneity and observer variation on the use of physical stream attributes as monitoring tools. For most stream habitat attributes evaluated, difference among streams accounted for greater than 80% of the total survey variation. To minimize the effect variation among streams has on evaluating stream conditions it may be necessary to design survey protocols and analysis that includes stratification, permanent sites, and/or analysis of covariance. Although total variation was primarily due to differences among streams, observers also differed in their evaluation of stream attributes. This study suggests if trained observers within a study designed to account for environmental heterogeneity evaluate objectively defined stream attributes, results should prove valuable in monitoring differences in reach-scale stream conditions. The failure to address any of these factors will likely lead to the failure of stream attributes as effective monitoring tools.

Key Terms; aquatic ecosystems, quality control, random effects model, sample size, <u>stream monitoring</u>, environmental heterogeneity.

Introduction

The last several decades have seen a proliferation of survey methods designed to describe and evaluate stream characteristics at the reach scale (Bain et al., 1999; Bauer and Ralph, 1999; Johnson et al., 2001). These protocols are often used to monitor anthropogenic-mediated changes in streams conditions (McDonald et al., 1991; Kaufmann et al., 1999). Evaluations of physical stream habitat frequently occur in lieu of monitoring biological conditions (i.e. fish density and biomass) because of the relative ease associated with collecting these data and the variability of biological systems (Platts et al., 1983). The outcomes of these surveys are used as evidence for compliance or noncompliance with federal laws (National Marine Fisheries Service, 1996; Bauer and Ralph, 2001), federal land management plans (Epsy and Babbit, 1994), and state laws (Washington State, 2001).

The use of stream attributes for monitoring has many critics. The most common critique has been that different observers using the same protocol often arrive at different results (Ralph et al., 1992; Roper and Scarnecchia, 1995; Poole et al., 1997). Additional concerns include inconsistent application of protocols (Hey and Thorne, 1983; Kondolf, 1997), lack of consistent training (Wang et al., 1996), and difficultly in using stream attributes to detect change due to management activity (McDonald et al., 1991; Kaufmann et al., 1999; Bauer and Ralph, 2001). At the root of each of these concerns is the variability associated with the measurement of an attribute and how it affects conclusions relative to that attribute.

Variability among how observers evaluate a stream metric has been used to justify discarding the metric in monitoring protocols (Roper and Scarnecchia, 1995; Poole et al., 1997; Bauer and Ralph, 2001). Variability, however, is a component of all sampling protocols (Ramsey et al., 1992; Clark et al., 1996) and is often an important aspect of understanding natural systems (Green, 1984; Frissel and Bayles, 1996). Therefore, instead of using variability as a rational for discarding an attribute, we suggest an understanding of variability be included in the design of a monitoring program and in the selection of attributes used in that program (Larson et al. 2001).

Variability in the measurement of stream attributes can be divided into three sources: environmental heterogeneity, sampling variance, and measurement error (Ramsey et al., 1992; Clark et al., 1996). For example, consider a study comparing median riffle particle size (D_{50}) between a set of streams where cattle are permitted to graze and a set of streams where cattle are excluded. The ability to discern differences will be affected by; 1) environmental heterogeneity due to the substrate differences among the streams surveyed, 2) sampling variance due to differences in how and where different observers pick up particles to be measured, and 3) the measurement error due to variability associated with how an observer measures an individual rock. Previous papers evaluating stream attributes have primarily addressed the observer (sampling variance + measurement error) portion of variability (Marcus et al. 1995; Wang et al., 1996; Poole et al., 1997; for an exception see Kaufmann et al., 1999).

The objective of this paper is to evaluate how variability in 13 common physical stream attributes affects their use in monitoring stream conditions. To meet this objective we evaluate the relative importance of variability among streams (environmental heterogeneity) and among observers (both sampling and measurement error), use estimates of total variation to determine sample sizes necessary to detect specified differences between group means of streams treated differently, and suggest variables likely to prove useful in monitoring changes in stream habitat.

Methods

Study Site

Data for this study were collected during the summer of 2000 on lands managed by the USDA Forest Service within Central Idaho, USA. A single reach in each of six different streams was surveyed. Stream reaches were chosen to represent a range of low gradient (<2% slope) wadeable streams within managed watersheds in Central Idaho. Within these constraints, streams varied in size, elevation, and geology (Table 1).

Study Design

Stream reaches were independently surveyed with either six or seven two-person crews. All crew members in this study received 10 days of training prior to completing surveys. Training consisted of collecting data while under the supervision of individuals who had designed the survey protocol. The intent of training was to minimize variability in how crews assessed stream attributes. Each two-person crew was responsible for estimates of the stream attributes evaluated for this study. Estimates of observer variability were based on the variability among these crews.

Test surveys were conducted during two one-week sessions: one in June and the other in August. Each session involved all crews evaluating a reach in three streams. The order crews sampled stream reaches was randomly assigned. Crews began surveying each creek in identical locations and sampled moving upstream a minimum distance of 20 times the bankfull width. Crews were asked to stop surveying at the bottom of the first pool they encountered after surveying the minimum distance. Although a pre-specified ending location could have been used, minor variation in where observers end surveys is an unavoidable aspect of most surveys so evaluation or these differences are included in the analysis (see Table 1). No precipitation fell during either session so stream behavior remained relatively consistent throughout each test period.

Physical Characteristics

Observers from these six/seven crews made independent observations relative to 13 different physical attributes within each reach. These attributes are commonly evaluated physical characteristics in stream monitoring protocols (Platts et al., 1987; MacDonald et al., 1991). The sampling methodologies originated from existing protocols. Some methodologies were refined following two years of use and evaluation (Archer et al., In Review). Additional changes were made to fit within the project study design of determining a stream reach average for each attribute. The original citation for each methodology and a brief description of the protocol is described below.

Gradient – Reach gradient was calculated by dividing the average elevation change (measured to the nearest cm) by reach length (measured to the nearest 0.1 m: Harrelson et al., 1994). Elevation change was estimated as the average of two independent measurements of elevation change using a 20X level-transit. Reach length was measured along the thalweg using a tape measure.

Sinuosity – Sinuosity was estimated by dividing the reach length by the straight-line distance of the reach. The straight-line distance was measured with a tape measure between the start and end of the reach. Both straight-line and reach length were measured to the nearest 0.1 m.

Streambank measurements were used to estimate a reach average for bank angle, undercut depth, percent undercut, and bank stability. Bank angle measurement techniques were taken from Platts et al. (1987) and bank stability from Bauer and Burton (1993). Measurements were taken from sample plots on both banks at a minimum of 20 equally spaced (every bankfull width) transects throughout the reach.

Bank Angle – The average bank angle was calculated as the average of the bank angles measured at each sample plot. Bank angles were measured by laying a depth rod along the bank at right angle to the flow. A clinometer was placed along the depth rod and the angle recorded to the nearest degree.

Percent Undercut – The percent of a reach with undercut banks was calculated by dividing the number of sample plots that were undercut by the total number of plots. Undercut banks had an angle < 90° and an undercut depth of > 5 cm.

Undercut Depth – The average undercut depth was calculated by summing the depths measured at each of the undercut bank locations and then dividing by the total number of sample plots. Undercut depth was measured to the nearest cm from the deepest part of the undercut to the outer edge of the streambank.

Percent Stable Banks – The percent of a reach with stable banks was calculated by dividing the number of sample plots with stable banks by the total number of plots. Each plot was consider stable if there was no sign of erosion, slumps, or fractures.

Bankfull Width – The bankfull width was calculated by averaging the width estimates from four cross-sections. The cross-sections were located at the widest point in four different fast-water habitat units and recorded to the nearest 0.1 m. Bankfull widths were estimated at fast-water habitat units instead of in a straight reach between two meander bends as suggested by Harrelson et al. (1994). This modification was made in an effort to increase the consistency in where crews took measurements. Bankfull indicators as described by Harrelson et al. (1994) were used to determine bankfull elevation.

Width-To-Depth Ratio – The average width-to-depth ratio was determined from the same four cross-sections used to determine bankfull width. The average depth at each cross-section was estimated from at least 10 evenly spaced bankfull depths instead of the 20 recommended by Harrelson et al. (1994). Fewer measurements were taken at each cross-section because our surveys were interested in width-to-depth ratios only, not accurately describing channel profiles. Depth was recorded to the nearest cm. The width-to-depth ratio at each cross-section was calculated as the bankfull width divided by the average depth. An overall reach average was calculated from these four cross-sections.

D_{50} – The D_{50} was the intermediate axis diameter of the median particle collected from particle counts (Wolman, 1954) within four fast-water habitat units within the reach. At least 25 bed-material particles were collected along transects within each unit for a minimum of 100 particles. Particles were measured to the nearest mm.

Percent Fines – The percent fines was the percent of the substrate collected for the estimation of D_{50} that had an intermediate axis diameter less than 6 mm.

Pool Tail fines – The percent of surface fine sediment was averaged from four pool tails. A 50-intersection grid (Bunte and Abt, 2001) was randomly tossed three times within each pool tail. The percent fines were calculated by dividing the number of intersections underlain with fine sediment (< 6 mm) by the total number of intersections (150).

Percent Pools – The percent of a reach comprised of pool habitat was estimated by summing the lengths of each pool and dividing by the reach length. Pool lengths were measured along the thalweg.

Residual Pool Depth – The average residual pool depth was calculated by averaging the individual residual depths from all pools encountered in the reach. Residual depth for each pool was derived by subtracting the pool tail depth from the maximum depth (Lisle, 1987). Depth at both pool tail and maximum were recorded to the nearest cm.

Analysis

A random effects analysis of variance model (PROC MIXED of SAS®) was used to estimate the means and variance associated with streams (stream variability) and crews (observer variability). Estimates of variance were evaluated by treating both stream and observer as a random effect within the model (Littell et al., 1996). We used a random effects model because we wished to generalize results to a larger population of streams and observers. All error in the model not associated with difference among streams was assumed to be due to differences among observers.

Estimates of variance for streams and observers were additive and can be used to calculate the proportion of the variation due to each component (Montgomery, 1984; Ramsey et al., 1992; Clark et al., 1996). Estimates of the variance and means were also used to calculate the coefficient of variation (CV, $(\text{variance})^{0.5}/\text{mean}*100$) for each variable.

We calculated minimum sample sizes necessary to detect differences between two population means (Zar, 1996) to provide guidance on which variables would be best used in monitoring differences in stream conditions among groups of streams treated differently (e.g., grazed vs. ungrazed). Sample size estimates are a good tool to evaluate monitoring attributes because they indicate the effort needed to be confident changes in an attribute will be detected (Eckblad, 1991; Hogle et al., 1993).

Minimum sample sizes were calculated using specified differences between two means, Type I, and Type II error rates. We evaluated differences between means that ranged from 5 to 50%. For this exercise the value of 5% was chosen for sampling practicality (Eckblad 1991). Variation in ecosystems is high so detecting differences less than 5% almost always require 100's if not 1000's of samples. The upper value, 50%, was chosen because when differences among groups are of this magnitude they have can have consequences to land management activities thought to influence them (Epsy and Babbit 1994, NMFS 1996; their Table 1).

In calculating sample sizes we were concerned with two types of statistical errors; 1) a Type I error (α), which is rejecting the null hypothesis (no mean difference between groups) when the null hypothesis is actually true and 2) a Type II error (β), which is failing to reject the null hypothesis when it is actually false (Peterman 1990). We limited our evaluation to a single Type I error rate, $\alpha = 0.1$, and two Type II error rates, $\beta = 0.1$ and $\beta = 0.25$. Type I errors are commonly reported in the literature. Knowing the Type II error rate helps ensure that findings of no significant difference between groups are not used as evidence that the null hypothesis is true unless the test also has high statistical power (Robinson and Wainer 2002).

Estimates of sample size were calculated following the iterative procedure outlined by Zar (1996, page 133, equation 8.22):

$$n \geq \frac{2\,s^2}{d^2}\,(t_{\alpha(2),\nu} + t_{\beta(1),\nu})^2$$

where s^2 is our estimate of total variance, $t_{\alpha(2),v}$ is the t-value for an acceptable two-tailed Type I error rate, $t_{\beta(1),\nu}$ is the t-value for an acceptable one tailed Type II error rate, ν is 2*(n-1) the degrees of freedom for s^2, and d is the minimum difference one wishes to detect. This equation calculates the number of samples needed from each population assuming equal sample sizes and equal variances. If the number of samples from one population is constrained (e.g. few unmanaged stream reaches), it is necessary to adjust the sample size of the unconstrained population (Zar, 1996). If variance of the two groups is unequal, an estimate of pooled variance should be used. When n exceeded 30, values for infinite sample size were substituted (z-values for t-values) because differences in results were minimal.

All analyses were conducted using raw data. Three variables, gradient, width-to-depth, and D_{50} were skewed but because of the small number of streams sampled, statistical models being robust to departures from normality (Zar, 1996), no underling theory suggesting data should be transformed (for discussion see Cohen and Cohen, 1983), and raw data resulting in conservative (larger) estimates of variance these attributes were not transformed. For example, gradient of the six streams we surveyed could be described as a log normal distribution (see Table 1). There is, however, no reason to think the distribution of all stream reaches between 0 and 2% gradient would have this same distribution. If the underlying distributions of these attributes are truly skewed, transforming data could reduce sample sizes needed to detect mean differences between groups.

Results

The results of this study suggest many commonly evaluated physical characteristics of streams can have considerable variation when measured and variation can be attributed to both differences among streams and differences in how observers evaluate streams. This variation has important implications on the design and interpretation of monitoring programs based on physical stream habitat variables.

Percent of Total Variation

In 10 of the 13 physical stream attributes evaluated, variation among streams accounted for 80% or more of the total variation (Table 2). In one variable, percent fines, stream variation accounted for about two-thirds of the total variation. Variation in two variables, percent stable banks and percent pools, was nearly equally divided between streams and observers.

Coefficient of Variation

Coefficients of variation (CV) for a given variable were considerably higher among streams than among observers. Estimates of variation resulted in the CV's ranging from 5 to 72 among streams and 5 to 38 among observers (Table 2). Coefficients of variation for all but three variables, bank stability, percent fines, and percent pool, were considerably greater among streams than among observers.

High CV's among streams indicated values of the measured attributes varied greatly among the sampled streams. Only percent stable banks showed little variation among the streams sampled (CV<10). Four variables, sinuosity, bank angle, percent pool, and residual pool depth, had CV's between 10 and 25. The remaining attributes all had stream CV's indicating average deviation was at least one-third the mean. The most variable attribute among streams was gradient. This was true even though we only evaluated stream with gradients less than 2%. Within this narrow stratum of streams, the standard deviation among streams still was nearly three-fourths the mean.

High CV's among observers indicated disagreement among how observers evaluated an attribute. When observer CV is zero, all observers evaluate the attribute in the same manner. Four variables, gradient, sinuosity, bank angle, and bank stability, had CV's that were not appreciably greater than zero (CV<10). Approximately one-half of the attributes had CV's between 10 and 25. All three substrate attributes, D_{50}, percent fines, and pool-tail fines, all had CV's in excess of 20.

Sample Size

Of the 13 variables evaluated using protocols specified in this study (Table 3), only percent stable banks could discern small changes (5%) samples sizes of less than 50 per group. If the desired minimum detectable difference between means was 10% and the Type II error rate increased to 0.25, sample sizes less than 65 from each group would be acceptable for five -- sinuosity, bank angle, bank stability, percent pools, and residual pool depth -- of the 13 variables. Four attributes, gradient, D_{50}, percent fines, and pool-tail fines, consistently required twice as many samples as the remaining attributes to detect the same magnitude of difference between groups.

Discussion

Most studies evaluating stream attributes as a monitoring tool have focused on variability associated with observers (Hogle et al., 1993; Roper and Scarnecchia, 1995;

Wang et al., 1996). These studies found high observer variability and suggested monitoring using stream attributes either be discontinued or include more operational definition of stream attributes and additional training of observers (Platts et al, 1983). This study suggest that while minimizing variability among observers is a necessary step in designing monitoring protocols (Peterson and Wolrab, 1999), it may be difficult to totally eliminate this source of variation. Even with the 10 days of training used in this study, the standard deviation for observers in 9 of the 13 variables exceed 10% of the mean and two of the variables has standard deviations exceeding 30% of the mean value.

With extensive training and protocol development we did find that for 10 of the 13 attributes observer variability accounted for 20% or less of the total variability. Studies evaluating water chemistry (Clark et al., 1996), geochemistry (Ramsey et al., 1992) and stream attributes (Kaufmann et al., 1999) have all suggested that when sampling variance (variation among observers; noise) accounts for 20% or less of the total variability it is likely to be a reliable monitoring attribute. Using this criterion, all but percent stable banks, percent fines, and percent pools should prove reliable attributes to measure when monitoring.

The problem with relying on the percent total variability attributable to differences among observers (or signal-to-noise ratio as proposed by Kaufmann et al., 1999) as the sole guide for selecting monitoring attributes is this measure fails to account for the total variance associated with the measurement of an attribute. For example, observer variability accounted for 20% of the total variability associated with evaluating pool tail fines. But the CV's associated with pool tail fines indicated the average deviation among observers was 34% of the mean value while the average deviation among streams was 69% of the mean value. Of all the attributes evaluated, this variable required the largest sample size to detect differences among sets of stream (Table 3). If one of the objectives of a monitoring program is to detect differences among treatments (when they exist) at the lowest cost, then both the total variance (sample size) and the proportion of the variability due to observers needs to be of concern.

Our data indicate that for most physical attributes, variability among streams has a far greater consequence on monitoring efforts than does differences among observers. The sample size equation is calculated using total variance (s^2_{total}), which is the sum of the variability due to streams (s^2_{stream}) and observers ($s^2_{observer}$). This means the proportion of the total sample due to stream and observer is directly related to its proportion of the total variability. This suggests that for all but three variables, bank stability, percent fines, and percent pools, the overwhelming majority (≥80%) of the samples needed in a monitoring program are driven by differences among streams not differences among observers.

Stream differences are a product of large-scale geomorphic features (Isaak, 2001; Jensen et al. 2001), the position of the stream within the watershed (Montgomery and Buffington 1997), effects of management (Ralph et al. 1993, Wood-Smith and Buffington 1996), and disturbance history (Reeves et al 1995). Monitoring efforts that rely on describing physical stream attributes must therefore focus additional effort on minimizing the effect variability among streams have on addressing the monitoring objective. Variation among streams can be reduced through stratification, measuring attributes at permanent sites, and/or by incorporating analysis of covariance into comparisons.

Stratification can appreciably reduce variation and result in more precise estimates of an attribute (Montgomery 1984). In a stream-monitoring program, stratification should focus on minimizing the effect stream formation processes have on differences in stream attributes. Landscape characteristics that may prove useful in stratifying include; geology (Hicks, 1989), watershed size (Lee et al., 1997), channel type (Rosgen, 1994; Montgomery and Buffington, 1997), and anthropogenic disturbance level (Dose and Roper, 1994; McIntosh et al., 2000). Characteristics used to stratify stream systems will depend upon the objective for which the data is collected. The primary disadvantage of stratification is that it can result in the number of sampling units for some strata being limited.

A second strategy for reducing variation among streams is sampling at permanent sites. The use of permanent sites has the potential to reduce the variance and therefore the number of samples needed when there is a high degree of correlation between sampling unit values through time (Elzinga et al., 1998). The disadvantage of permanent sites is that more effort is focused in fewer locations so results may have limited applicability towards answering larger scale questions.

A third strategy for reducing variation among streams is using analysis of covariance (ANCOVA). The use of ANCOVA removes variability in an experimental unit that could not be controlled for in the survey design (Littell et al., 1996). This is a common occurrence in studies on the effects of land management. For example, managed watersheds can differ from unmanaged watersheds in their gradient, basin area, and elevation. Because these differences are difficult to account for through stratification (i.e., few large, low-elevation unmanaged watershed), ANCOVA may be the only tool to reduce the amount of variability among streams. ANCOVA is best applied to variables unaffected by the process under study such as gradient, geology, or elevation in the case of evaluating the effects of land management. Inferences based on ANCOVA can be difficult to interpret if there is considerable error in measuring the attribute used as the covariate or when there are sizeable differences between means of the evaluated groups (Dunn and Clark, 1987).

Although total variation, how it is partitioned between streams and observers, and CV's, all provide important information necessary in selecting an attribute to be evaluated, it cannot be the only justification for deciding which stream attributes to monitor. Selection of variables must also include an understanding of how relevant the attribute is to the biota of concern, how rapidly the attribute can change, and how land management might alter the attribute (Robison, 1997; Bauer and Ralph, 2001, Montgomery and MacDonald 2002).

Of the attributes we evaluated, only those associated with describing the stream substrate (D_{50}, percent fines, pool tail fines) may prove impractical in evaluating stream conditions subject to different treatments. All three of these attributes had high total variation resulting in much higher total sample sizes than most other variables (Table 3). Our study suggests alternative protocols need to be developed to more precisely describe stream substrate at the reach level (Kondolf, 1997; Bunte and Abt, 2001).

Conclusion

The primary objectives of this study was evaluate variation associated with the measurement of physical stream characteristics and suggest variables that would be useful in monitoring physical stream habitat characteristics. Our evaluation of these 13 attributes suggests that given extensive training and protocol development, all but three

(D_{50}, percent fines and pool tail fines) would likely prove useful in discerning differences in streams subject to different treatments.

Due to high total variance, most attributes are best suited for well designed well-funded monitoring programs. In contrast, these attributes are ill suited for the quick evaluation of single streams by poorly trained observers. The success of monitoring studies between these two extremes will be based on formulating specific monitoring objectives (Kershner et al., 1991), careful selection of stream attributes based on the estimates of variation found in this study, pilot studies conducted in the region of interest, an understanding of the magnitude of change the attribute is likely to exhibit (Montgomery and MacDonald 2002), and sufficient resources to conduct the monitoring study as planned.

Even in well-designed studies, monitoring objectives will have to be formulated in a way that provides context for both statistical and biological significance. The variance associated with the measurement of these attributes could lead to cases where there is a failure to find statistically significant differences ($\alpha<0.5$) between groups of streams (i.e. managed vs. unmanaged), yet the means of the two groups differ by 10 to 30%. Interpretation of such results cannot be based on statistical significance alone, but must include an understanding of monitoring objectives, biological significance, and the cost of Type I and Type II errors (Peterman, 1990; Johnson, 1999; Anderson et al., 2000).

Our results suggest careful evaluation of variability associated with the measurement of stream attributes can improve sampling designs so that many physical stream attributes can be used to monitor stream conditions. The high total variability associated with the measurement of these variables suggest failure to incorporate a well designed sampling scheme, extensive training, and operational protocols will result in stream monitoring programs that fails to detect altered or changing stream conditions.

Acknowledgements

We would like to thank the summer field technicians who through a combination of hard work and patience collecting the data on which this analysis is based. Paul Badame, Chad Mellison, and Dave Menea insured data integrity as data collected in the field was summarized for analysis. This paper was greatly improved by John Potyondy, Dan Isaak and three anonymous reviewers. We received statistical guidance from Dave Turner, Richard Cutler, and Susan Durham. Region's 1,4, and 6 of the USDA Forest Service and USDI Bureau of Land Management's state offices of Oregon/Washington and Idaho provided funding for this project.

Literature Cited

Anderson, D.R., and K.P. Burnham, and W.L. Thompson. 2000. Null Hypothesis Testing, Prevalence, and an Alternative. Journal of Wildlife Management 64:912-923.

Archer, E.K., S.C. Mellison, R. Henserson, J.L Kershner, and B.B. Roper. In Review. Testing Common Stream Sampling Methods: How Useful are these Techniques for Broad-Scale, Long-Term Monitoring? USDA Forest Service, Rocky Mountain Research Station General Technical Report

Bain, M.B., T.C. Hughes, and K.K. Arend. 1999. Trends in Methods for Assessing Freshwater Habitats. Fisheries 24(4)16-21.

Bauer S.B. and T.A. Burton. 1993. Monitoring Protocols to Evaluate Water Quality Effects of Grazing Management on Western Rangeland Streams. Environmental Protection Agency, Region 10, EPA 910/R-93-017, Seattle, WA

Bauer, S.B. and S.C. Ralph. 1999. Aquatic Habitat Indicators and their Application to Water Quality Objectives within the Clean Water Act. EPA-910-R-99-014. US Environmental Protection Agency, Region 10, Seattle WA.

Bauer, S.B and S.C. Ralph. 2001. Strengthening the Use of Aquatic Habitat Indicators in the Clean Water Act Programs. Fisheries 26(6):14-25.

Bunte, K. and S.R. Abt. 2001. Sampling Frame for Improving Count Accuracy in Coarse Gravel-Bed Streams. Journal of the American Water Resources Association 37:1001-1014.

Clark, M.J., MCA Laidlaw, S.C. Ryneveld, and M.I. Ward. 1996. Estimating Sampling Variance and Local Environmental Heterogeneity for both Known and Estimated Analytical Variance. Chemosphere 32:1133-1151.

Cohen, J. and P Cohen. 1983. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (second edition). Lawrence Erlbaum Associates, Publishers. Hillside N.J. 545 pgs.

Dose, J.J., and B.B. Roper. 1994. Long-Term Changes in Low-Flow Channel Widths within the South Umpqua Watershed, Oregon. Water Resources Bulletin 30:99-1000.

Dunn, O.J. and V.A. Clark. 1987. Applied Statistics: Analysis of Variance and Regression. John Wiley and Sons, NY, NY. 445 pgs.

Eckblad, J.W. 1991. How Many Samples Should be Taken. BioScience 41:346-348.

Elzinga, C.L., D.W. Salzer, J.W. Willoughby. 1998. Measuring and Monitoring Plant Populations. BLM Technical Reference 1730-1. BLM (Bureau of Land Management)/RS/ST-98/005+1730

Espy, M. and B. Babbit. 1994. Record of Decision for Amendments to the Forest Service and Bureau of Land Management Planning Documents within the Range of the Northern Spotted Owl. 1994-589-111/00001. United States Printing Office, Washington D.C.

Frissell, C.A. and D. Bayles. 1996. Ecosystem Management and the Conservation of Aquatic Biodiversity and Ecological Integrity. Water Resources Bulletin 32:229-240.

Green, R.H. 1984. Statistical and Nonstatistical Considerations for Environmental Monitoring Studies. Environmental Monitoring and Assessment 4:293-301.

Harrelson, C.C., C.L. Rawlins, and J.P. Potyondy. 1994. Stream Channel Reference Sites: and Illustrated Guide to Field Technique. General Technical Report RM-245. Fort Collins, CO; US Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 61 p.

Hey, R.D. and C.R. Thorne. 1983. Accuracy of Surface Samples from Gravel Bed Material. Journal of Hydraulic Engineering 109:842-851.

Hicks, B.J. 1989. The Influence of Geology and Timber Harvest on Channel Morphology and Salmonid Populations in Oregon Coast Range streams. PhD. Dissertation. Oregon State University, Corvallis OR 155 pages.

Holge, J.S., T.A. Wesche, and W. A. Hubert. 1993. A Test of the Precision of the Habitat Quality Index Model II. North American Journal of Fisheries Management 13:640-643.

Isaak, Daniel J. 2001. A Landscape Ecological View of Trout Populations across a Rocky Mountain Watershed. PhD. Dissertation. University of Wyoming, Laramie WY. 148 pages.

Jensen, M.E., I.A. Goodman, P.S. Bourgeron, N.L. Poff, and C.K. Brewer. 2001. Effectiveness of Biophysical Criteria in the Hierarchical Classification of Drainage Basins. Journal of the American Water Resources Association 37:1155-1167.

Johnson, D.H. 1999. The Insignificance of Statistical Significance Testing. Journal of Wildlife Management 63:763-772.

Johnson, D.H., N. Pittman, E. Wilder, J.A. Silver, R.W. Plotnikoff, B.C. Mason, K.K. Jones, P. Roger, T.A. O'Neil, C. Barrett. 2001. Inventory and Monitoring of Salmon Habitat in the Pacific Northwest – Directory and Synthesis of Protocols for Management/Research and Volunteers in Washington, Oregon, Idaho, Montana, and British Columbia. Washington Department of Fish and Wildlife, Olympia Washington. 212 pages.

Kaufmann, P.R., P. Levine, E. G. Robison, C. Seeliger, and D.V. Peck. 1999. Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003 U.S. Environmental Protection Agency, Washington D.C.

Kershner, J.L., H.L. Forsgren, and W.R. Meehan. 1991. Managing Salmonid Habitats. American Fisheries Society Special Publication 19:599-606.

Kondolf, G.M. 1997. Application of Pebble Count: Notes on Purpose, Method, and Variants. Journal of the American Water Resources Association 33:1395-1399.

Larsen, D.P., T.M. Kincaid, S.E. Jacobs, and N.S. Urquhaart. 2001. Designs for evaluating local and regional trends. Bioscience 51:1069-1078.

Lee, D.C. J.R. Sedell, B.E. Rieman, R.F. Thurow, and J.E. Williams. 1997. An Assessment of Ecosystem Components in the Interior Columbia Basin and Portions of the Klamath and Great Basin. Volume 3. Broadscale Assessment of Aquatic Species and Habitats. USDA Forest Service General Technical Report PNW-GTR-405. Portland Oregon.

Lisle, T. E. 1987. Using "Residual Depths" to Monitor Pool Depths Independently of Discharge. Research Note PSW-394, USDA, Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California.

Littell, R.C., G.A. Miliken, W.W. Stroup, R.D. Wolfinger. 1996. SAS® System for Mixed Models. SAS Institute Inc. Cary, NC. 633 pgs

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streamin the Pacific Northwest and Alaska. EPA-910/9-001. US Environmental Protection Agency, Region 10, Seattle WA.

Marcus, W.A., S.C. Ladd, J.A. Stroughton, and J.W. Stock. 1995. Pebble counts and the role of user-dependent bias in documenting sediment size distributions. Water Resources Research 31:2625-2631.

McIntosh, B.A., J.R. Sedell, R.F. Thurow, S.E. Clark, and G.L. Chandler. 2000. Historical Changes in Pool Habitats in the Columbia River Basin. Ecological Applications 10:1478-1496.

Montgomery, D.C. 1984. Design and Analysis of Experiments. John Wiley and Sons, NY, NY.

Montgomery, D.R. and J.M. Buffington. 1997. Channel-Reach Morphology in Mountain Drainage Basins. Geological Society of America Bulletin 109:596-611.

Montgomery, D.R., and L.H. MacDonald. 2002. Diagnostic approach to stream channel assessment and monitoring. Journal of the American Water Resources Association 38:1-16.

National Marine Fisheries Service. 1996. Making Endangered Species Act Determinations of Effects for Individual or Grouped Actions at the Watershed Scale. National Marine Fisheries Service, Environmental and Technical Service Division. 32 pgs. (electronic document at www.nwr.noaa.gov/1habcon/habweb/pubs/matrix.pdf)

Peterman, R.M. 1990. Statistical Power can Improve Fisheries Research and Management. Canadian Journal of Fisheries and Aquatic Sciences 47:2-15.

Peterson, James T. and S. P. Wollrab. 1999. An Analysis of Potential Stream Fish and Fish Habitat Monitoring Procedures for the Inland Northwest. Annual Report 1999 to Bonneville Power Administration, Portland, OR, Contract No. 92AI25866, Project No. 92-032-00, 61 electronic pages (BPA Report DOE/BP-25866-1)

Platts, W.S., W.F. Megahan, G.W. Minshall. 1983. Methods for Evaluating Stream, Riparian, and Biotic Conditions. US Department of Agriculture, Forest Service. General Technical Report INT-138 Ogden Ut. 70 pages.

Platts, W.S., C. Armour, G.D. Booth, M. Bryant, J.L. Bufford, P. Cuplin, S. Jensen, G. W. Lienkaemper, G.W. Minshall, S.P. Monsen, R.L. Nelson, J.R. Sedell, amd J.S. Tuhy. 1987. Methods for Evaluating Riparian Habitats with Applications to Management. Intermountain Research Station General Technical Report INT-221. 177 pages.

Poole, G.C., C.A Frissell, and S.C. Ralph. 1997. In-Stream Habitat Unit Classification: Inadequacies for Monitoring and Some Consequences for Management. Journal of the American Water Resources Association 33:879-896.

Ralph, S.C., T Cardoso, C.G. Poole, L.C. Conquest, and R.J. Naiman. 1992. Status and Trends of Instream Habitat in Forested Lands of Washington: The Timber, Fish, and Wildlife Ambient Monitoring Project – 1989-1991. Biennial Progress Report. University of Washington, Center for Streamside Studies Report to the Washington Department of Natural Resources, Olympia WA.

Ralph, S.C., G. C. Poole, L.L. Conquest, and R.J. Naiman. 1993. Stream channel morphology and woody debris in logged and unlogged basins of Western Washington. Canadian Journal of Fish and Aquatic Sciences 51:37-51

Ramsey, M.H., M Thompson, and M. Hale. 1992. Objective Evaluation of Precision Requirements for Geochemistey Analysis using Robust Analysis of Variance. Journal of Geochemical Exploration 44:23-36.

Reeves, G.H., L.E. Benda, K.M. Burnett, P.A. Bisson, and J.R.Sedell. 1995. A Disturbance-Based Ecosystem Approach to Maintaining and Restoring Freshwater Habitats of Evolutionarily Significant Units of Anadromous Salmonids in the Pacific Northwest. American Fisheries Society Symposium 17:334-349.

Robinson, D.H. and H. Wainer. 2002. On the past and future of null hypothesis significance testing. Journal of Wildlife Management 66:263-271.

Robison E.G. 1997. Reach Scale Sampling Metrics and Longitudinal Pattern Adjustment of Small Streams. Doctoral Dissertation in Forest Engineering and Civil, Construction, and Environmental Engineering. Oregon State University, Corvallis OR 254 pages.

Roper, B.B. and D.L. Scarnecchia. 1995. Observer Variability in Classifying Habitat Types in Stream Surveys. North American Journal of Fisheries Management 15:49-53.

Rosgen, D.L. 1994. A Classification of Natural Rivers. Catena 22:169-199.

Wang, L., T.D. Simonson, and J. Lyons. 1996. Accuracy and Precision of Selected Stream Habitat Estimates. North American Journal of Fisheries Management 16:340-347.

Washington State. 2001. Washington Forest Practice Rule Book (Title 222 WAC). Department of Natural Resources, Forest Practices Division. Olympia WA. 268 pgs. (www.wa.gov/dnr/htdocs/fp/fpb/fprules2001/wac222final.pdf)

Wolman, M.G. 1954. A Method of Sampling Coarse River-Bed Material. Transactions of the American Geophysical Union 35:951-956.

Wood-Smith, R.D. and J.M. Buffington. 1996. Multivariate geomorphic analysis of forest streams: implications for assessment of land use impacts on channel conditions. Earth Surface Processes and Landforms 21:377-393.

Zar, J.H. 1996. Biological Analysis (third edition). Prentice Hall, Englewood Cliffs, N.J. 662 pgs.

Table 1. General characteristics of streams evaluated for this paper.

				Bankfull	Reach Length	
Stream	**Geology**	**Elevation (m)**	**Gradient**	**Width (m)**	**Min (m)**	**Max (m)**
Big	Granitic	1939	0.30	4.9	127.9	133.5
Jack	Granitic	1610	0.71	2.6	60.6	88.6
Meadow	Granitic	976	0.37	7.8	120.2	128.0
Boulder	Volcanic	1448	0.37	7.3	162.0	180.0
Little Goose	Volcanic	1524	1.58	3.9	84.0	91.9
Lost	Volcanic	1480	0.65	5.4	119.6	128.0

Table 2. Means, standard deviations (stdev), percent of variation (% Var), and coefficient of variation (CV) associated with the 13 physical stream attributes measured in this study.

	Total	Stream		Observer	
Stream Attribute	**Mean (stdev)**	**% Var**	**(CV)**	**% Var**	**(CV)**
Gradient (%)	0.66 (0.4811)	99	(72)	1	(8)
Sinuosity	1.95 (0.3303)	86	(16)	14	(6)
Bank Angle (°)	96.44 (21.0778)	85	(20)	15	(8)
Undercut Depth (m)	0.10 (0.0446)	83	(39)	17	(18)
Bank Undercut (%)	40.92 (19.5487)	86	(44)	14	(18)
Bank Stability (%)	91.80 (6.5506)	53	(5)	47	(5)
Bankfull Width (m)	5.30 (2.0768)	88	(37)	12	(13)
Width:Depth Ratio	14.57 (6.2364)	85	(40)	15	(16)
D_{50} (mm)	25.03 (16.7820)	89	(63)	11	(22)
Percent Fines	31.61 (21.4809)	69	(57)	31	(38)
Pool Tail Fines (%)	26.16 (20.0107)	80	(69)	20	(34)
Percent Pools	69.98 (13.1694)	44	(13)	56	(14)
Residual Pool Depth (m)	0.37 (0.0905)	80	(22)	20	(11)

Table 3. Minimum sample sizes needed to detect difference among physical variables when both observer variability and stream heterogeneity are accounted for. Sample size estimates assume equal sample size and variance (Zar 1996). Values listed below indicate half the total sample needed. The first value listed in each column is the sample size needed to detect the stated change with a Type I error of 0.1 and a Type II error of 0.1. The value in parentheses has the same Type I error rate but a Type II error rate of 0.25.

	Sample Sizes Needed to Detect Changes of:				
Stream Attribute	5%	10%	20%	30%	50%
Gradient (%)	3,611 (2,268)	903 (567)	226 (142)	101 (63)	37 (24)
Sinuosity	198 (125)	50 (32)	14 (9)	7 (5)	3 (2)
Bank Angle (°)	328 (206)	82 (52)	22 (14)	10 (7)	5 (3)
Undercut Depth (m)	1,282 (806)	321 (202)	81 (51)	36 (24)	15 (9)
Bank Undercut (%)	1,564 (938)	391 (246)	98 (62)	44 (29)	17 (11)
Bank Stability (%)	35 (23)	10 (7)	4 (3)	3 (2)	2 (2)
Bankfull Width (m)	1,054 (662)	264 (166)	66 (42)	30 (20)	12 (8)
Width: Depth Ratio	1,266 (789)	314 (198)	79 (50)	35 (23)	14 (9)
D_{50} (mm)	3,081 (1,935)	771 (483)	193 (121)	86 (54)	31 (21)
Percent Fines	3,165 (1,988)	792 (497)	198 (125)	88 (56)	32 (20)
Pool Tail Fines	4,011 (2,519)	1,003 (630)	251 (158)	112 (70)	41 (26)
Percent Pools	243 (153)	61 (39)	17 (11)	8 (6)	4 (3)
Residual Pool Depth (m)	405 (255)	102 (64)	27 (17)	13 (8)	6 (4)